On hitting times and fastest strong stationary times for skip-free and more general chains

JAMES ALLEN FILL[1]

Department of Applied Mathematics and Statistics

The Johns Hopkins University

jimfill@jhu.edu and http://www.ams.jhu.edu/~fill/

ABSTRACT

An (upward) skip-free Markov chain with the set of nonnegative integers as state space is a chain for which upward jumps may be only of unit size; there is no restriction on downward jumps. In a 1987 paper, Brown and Shao determined, for an irreducible continuous-time skip-free chain and any d, the passage time distribution from state 0 to state d. When the nonzero eigenvalues ν_j of the generator on $\{0, \ldots, d\}$, with d made absorbing, are all real, their result states that the passage time is distributed as the sum of d independent exponential random variables with rates ν_j. We give another proof of their theorem. In the case of birth-and-death chains, our proof leads to an explicit representation of the passage time as a sum of independent exponential random variables. Diaconis and Miclo recently obtained the first such representation, but our construction is much simpler.

We obtain similar (and new) results for a fastest strong stationary time T of an ergodic continuous-time skip-free chain with stochastically monotone time-reversal started in state 0, and we also obtain discrete-time analogs of all our results.

In the paper's final section we present extensions of our results to more general chains.

---

1   Introduction and summary

An (upward) skip-free Markov chain on the set of nonnegative integers is a chain for which upward jumps may be only of unit size; there is no restriction on downward jumps. Brown and Shao [1] determined, for an irreducible continuous-time skip-free chain and any d, the passage time distribution from state 0 to state d (we have equivalently re-identified the exponential rates):

Theorem 1.1 ([1]). *Consider an irreducible continuous-time skip-free chain Y on the nonnegative integers with $Y(0) = 0$. Given d, let X (with state space $\{0, \ldots, d\}$) be obtained from Y by making d an absorbing state, and let G denote the generator for X. Then the hitting time of state d (same for X and Y) has Laplace transform*

$$u \mapsto \prod_{j=0}^{d-1} \frac{\nu_j}{\nu_j + u},$$

where $\nu_0, \ldots, \nu_{d-1}$ are the d nonzero eigenvalues of $-G$ (known to have positive real parts). In particular, if the ν_j's are real, then the hitting time distribution is the convolution of $\mathrm{Exponential}(\nu_j)$ distributions.

In this paper we give a new proof for Theorem 1.1. In the case of birth-and-death chains (which are time-reversible and thus have real eigenvalues), the result is often attributed to Keilson but in fact (consult [3]) dates back at least to Karlin and Mc-Gregor [9]. Our proof leads in this case to an explicit (sample-path) representation of the passage time as a sum of independent exponential random variables. Diaconis and Miclo [3] recently obtained the first such representation for birth-and-death chains, but our construction is much simpler.

There is an analog for discrete time:

Theorem 1.2. *Consider an irreducible discrete-time skip-free chain Y on the nonnegative integers with $Y(0) = 0$. Given d, let X (with state space $\{0, \ldots, d\}$) be obtained from Y by making d an absorbing state, and let P denote the transition matrix for X. Then the hitting time of state d (same for X and Y) has probability generating function*

$$u \mapsto \prod_{j=0}^{d-1} \left[\frac{(1 - \theta_j)u}{1 - \theta_j u} \right],$$

where $\theta_0, \ldots, \theta_{d-1}$ are the d non-unit eigenvalues of P. In particular, if every θ_j is real and nonnegative, then the hitting time distribution is the convolution of $\mathrm{Geometric}(1 - \theta_j)$ distributions.

Remark 1.3. The conclusions of Theorems 1.1 and 1.2 hold for all skip-free chains X on $\{0, \ldots, d\}$ for which $g_{i,i+1} > 0$ (respectively, $p_{i,i+1} > 0$) for $i = 0, \ldots, d-1$. Indeed, our proofs extend to that case, or one can extend the theorems by a simple perturbation argument.

Results similar to Theorems 1.1–1.2 can also be established for fastest strong stationary times (consult [2], [4], and [5] for general background, and (3.3) and (3.5) of [5] for how to check from the generator G of a continuous-time chain whether the chain has the "monotone likelihood ratio" property, that is, whether its time-reversal is stochastically monotone):

Theorem 1.4.
(a) *Consider an ergodic (equivalently, irreducible) continuous-time skip-free chain X on the state space $\{0,\ldots,d\}$ with $X(0) = 0$ and stochastically monotone time-reversal. Let G denote the generator for X. Then a fastest strong stationary time (SST) for X has Laplace transform*

$$u \mapsto \prod_{j=0}^{d-1} \frac{\nu_j}{\nu_j + u},$$

where $\nu_0, \ldots, \nu_{d-1}$ are the d nonzero eigenvalues of $-G$ (known to have positive real parts). In particular, if the ν_j's are real, then the fastest SST distribution is the convolution of Exponential(ν_j) *distributions.*

(b) *Consider an ergodic (equivalently, irreducible and aperiodic) discrete-time skip-free chain X on the state space $\{0,\ldots,d\}$ with $X(0) = 0$ and stochastically monotone time-reversal. Let P denote the transition matrix for X. Then a fastest strong stationary time (SST) for X has probability generating function*

$$u \mapsto \prod_{j=0}^{d-1} \left[\frac{(1-\theta_j)u}{1 - \theta_j u} \right],$$

where $\theta_0, \ldots, \theta_{d-1}$ are the d non-unit eigenvalues of P. In particular, if every θ_j is real and nonnegative, then the fastest SST distribution is the convolution of Geometric($1 - \theta_j$) *distributions.*

Our proofs of Theorems 1.1–1.2 and 1.4 could hardly be simpler in concept. Take Theorem 1.2, for instance. Let $\theta_0, \ldots, \theta_{d-1}$ be ordered arbitrarily. We will exhibit a matrix Λ, with rows and columns indexed by the state space $\{0,\ldots,d\}$, having the following properties:

(i) Λ is lower triangular.

(ii) The rows of Λ sum to unity.

(iii) $\Lambda P = \widehat{P}\Lambda$, where $\widehat{P}$ is defined by

$$\hat{p}_{ij} := \begin{cases} \theta_i & \text{if } j = i \\ 1 - \theta_i & \text{if } j = i + 1 \\ 0 & \text{otherwise,} \end{cases} \tag{1.1}$$

for $i, j = 0, \ldots, d$ (with $\theta_d = 1$).

We will prove that from these three properties follows

$$\mathbf{P}(T \le t) = \widehat{P}^t(0, d), \quad t = 0, 1, \ldots, \tag{1.2}$$

where T is the absorption time in question. [In fact, all we will need to prove (1.2) are (iii) and the following two easy consequences of (i)–(ii):

(iv) $\Lambda(0, k) = \delta_{0,k}$.

(v) $\Lambda(k, d) = 0$ if $k \ne d$.]

In the special case that the θ_j's are real and nonnegative, then Theorem 1.2 follows immediately from (1.2); the general case will not take too much additional work. Moreover, for birth-and-death chains with nonnegative eigenvalues the proof will reveal how to identify (i.e., construct) geometric random variables summing to T.

In Section 1 of [6] it was shown how that paper's main continuous-time theorem (Theorem 1.1) follows immediately from the main discrete-time theorem (Theorem 1.2). By the same one-line proof, in the present paper Theorem 1.1 follows immediately from Theorem 1.2. For this reason we choose not to present a direct proof of Theorem 1.1, but in the case of birth-and-death chains we will give the explicit construction of independent exponential random variables summing to the absorption time. For the same reason we will prove only part (b) of Theorem 1.4.

Here is an outline for the paper. In Section 2 we prove Theorem 1.2, and in Section 3 we give stochastic constructions for Theorems 1.1–1.2 in the special case of birth-and-death chains. Section 4 is devoted to a proof of Theorem 1.4(b). In Section 5 we present extensions of our results, including stochastic constructions, to more general chains.

2 Proof of Theorem 1.2

In this section we prove Theorem 1.2, following the outline near the end of Section 1. Denote the eigenvalues of P, with each distinct eigenvalue listed as many times as its algebraic multiplicity, by $\theta_0, \ldots, \theta_d$. We claim that precisely one of these, say θ_d, equals 1 and that the rest have modulus smaller than 1; how the rest are ordered will for the majority of our results be irrelevant (we shall take special notice otherwise).

Here is a proof of the simple claim. Let P' denote the leading principal d-by-d submatrix of P. Expanding the determinant of $P - \theta I$ on the last row, we find that the $d + 1$ eigenvalues of P are $\theta = 1$ together with the d eigenvalues of P'. But by our assumptions that $p_{i,i+1} > 0$ for $t = 0, \ldots, d - 1$ and d is absorbing, we have $(P')^t \to 0$ and so (see for example [8, Theorem 5.6.12]) $\rho(P') < 1$.

Let I denote the identity matrix and define

$$Q_k := (1 - \theta_0)^{-1} \cdots (1 - \theta_{k-1})^{-1}(P - \theta_0 I) \cdots (P - \theta_{k-1}I), \quad k = 0, \ldots, d, \tag{2.1}$$

with the natural convention $Q_0 := I$. Note that

$$Q_k P = \theta_k Q_k + (1 - \theta_k)Q_{k+1}, \quad k = 0, \ldots, d - 1, \tag{2.2}$$

and that the rows of each Q_k all sum to 1.

Define
$$\Lambda(i,j) := Q_i(0,j), \quad i,j = 0,\dots,d. \tag{2.3}$$

Lemma 2.1. *Let P be as in Theorem 1.2. Then the matrix Λ defined at* (2.3) *enjoys properties* (i)–(iii) *listed in Section 1.*

Proof. The rows of Λ sum to unity because each of the basic factors $(1-\theta_r)^{-1}(P-\theta_r I)$ in (2.1) has that property; and since P is skip-free, Λ is lower triangular. Our next claim is that $\Lambda P = \widehat{P}\Lambda$, where $\widehat{P}$ is defined at (1.1). Indeed, equality of kth rows is clear from (2.2) for $k = 0,\dots,d-1$ and from the Cayley–Hamilton theorem for $k = d$.

Remark 2.2. It is not surprising that the eigenvalues of P line the diagonal of the upper triangular matrix $\widehat{P}$, because our Λ is nonsingular and so property (iii) implies that P and $\widehat{P}$ are similar. To see that the lower triangular matrix Λ is nonsingular, we check that its diagonal entries are all nonzero. Indeed, for $k = 0,\dots,d$, we use the skip-free property of P to calculate:

$$\Lambda(k,k) = Q_k(0,k) = \frac{p_{0,1}\cdots p_{k-1,k}}{(1-\theta_0)\cdots(1-\theta_{k-1})} \neq 0.$$

So we have found a matrix Λ satisfying properties (i)–(iii), and properties (iv)–(v) follow easily from (i)–(ii). Notice that property (iii) immediately extends to

$$\Lambda P^t = \widehat{P}^t \Lambda, \quad t = 0, 1, \dots. \tag{2.4}$$

Lemma 2.3. *Let P and X, with absorption time T, be as in Theorem 1.2 and let $\widehat{P}$ be the bidiagonal matrix defined at* (1.1). *Then* (1.2) *holds:*

$$\mathbf{P}(T \le t) = \widehat{P}^t(0,d), \quad t = 0, 1, \dots.$$

Proof. Let Λ be any matrix enjoying properties (iii)–(v) of Section 1; recall that Lemma 2.1 provides such a Λ. We will use Λ to establish the desired result, and the consequence (2.4) of property (iii) will be key.

We consider the $(0,d)$ entry of each side of (2.4). On the left we have

$$(\Lambda P^t)(0,d) = \sum_k \Lambda(0,k) P^t(k,d) = P^t(0,d) = \mathbf{P}(T \le t),$$

where the second equality is immediate from property (iv). On the right we have

$$(\widehat{P}^t \Lambda)(0,d) = \sum_k \widehat{P}^t(0,k)\Lambda(k,d) = \widehat{P}^t(0,d)\Lambda(d,d),$$

where the second equality is immediate from property (v). So all that remains is to establish that $\Lambda(d,d) = 1$.

We do this by passing to the limit as $t \to \infty$ in our now-established equation

$$\mathbf{P}(T \le t) = \widehat{P}^t(0,d)\Lambda(d,d).$$

The limit on the left is, of course, 1. Moreover, we claim that $\widehat{P}^t(0,d) \to 1$, and then the proof of the lemma will be complete. The claim is probabilistically obvious when the eigenvalues are all real and nonnegative. To see the claim in general, write $\widehat{P}$ as the bordered matrix

$$\widehat{P} = \left[\begin{array}{cc} A & b \\ 0 & 1 \end{array} \right]$$

by breaking off the last row and column. The matrix A has spectral radius $\rho(A) < 1$; and because the rows of $\widehat{P}$ sum to unity, we have $b = (I - A)\mathbf{1}$, where $\mathbf{1}$ denotes the vector of d ones. In this notation we find

$$\widehat{P}^t = \left[\begin{array}{cc} A^t & (I + A + \cdots + A^{t-1})b \\ 0 & 1 \end{array} \right] \to \left[\begin{array}{cc} 0 & (I - A)^{-1}b \\ 0 & 1 \end{array} \right],$$

which is a matrix with 1s throughout the last column and 0s elsewhere. In particular, $\widehat{P}^t(0,d) \to 1$. $\qquad\square$

As mentioned in Section 1, Lemma 2.3 is clearly all that is needed to prove Theorem 1.2 when the eigenvalues of P are nonnegative real numbers. In general we proceed as follows. From Lemma 2.3 it follows that T has probability generating function

$$\mathbf{E}\, u^T = (1 - u)\,(I - u\widehat{P})^{-1}(0,d), \quad |u| < 1. \tag{2.5}$$

But the simple form of $\widehat{P}$ makes it an easy matter to invert $1 - u\widehat{P}$ explicitly: the inverse is upper triangular, with

$$(I - u\widehat{P})^{-1}(i,j) = \frac{(1 - \theta_i) \cdots (1 - \theta_{j-1})u^{j-i}}{(1 - \theta_i u) \cdots (1 - \theta_j u)}, \qquad 0 \le i \le j \le d. \tag{2.6}$$

Taking $i = 0$ and $j = d$ and combining (2.5)–(2.6) gives Theorem 1.2.

3 Stochastic constructions for Theorems 1.1 and 1.2

In this section we consider Theorems 1.1 and 1.2 again in the case of a birth-and-death chain. For such a chain, the eigenvalues ν_j or θ_j are all real; this can be seen by perturbing (by an arbitrarily small amount) to get an ergodic generator or kernel, which is time-reversible and thus diagonally similar to a symmetric matrix. We show how the proof in Section 2 of Theorem 1.2 yields an explicit construction of independent geometric random variables summing to the hitting time in the case of a birth-and-death chain having nonnegative eigenvalues (for which it is sufficient that the holding probabilities satisfy $p_{ii} \ge 1/2$ for all i). We also exhibit an analogous representation of the hitting time in Theorem 1.1 as a sum of independent exponential random variables in the case of a continuous-time birth-and-death chain.

We begin with the discrete case. Returning to (2.1), we now suppose that P is a discrete-time birth-and-death kernel with nonnegative eigenvalues θ_j ordered so that $0 \le \theta_0 \le \cdots \le \theta_{d-1} < \theta_d = 1$. The polynomials $(P - \theta_0 I) \cdots (P - \theta_{k-1}I)$ in P appearing at (2.1) are called *spectral polynomials*. We claim that the spectral polynomials are all nonnegative matrices, that is, that the matrices Q_k at (2.1) are stochastic; this

follows immediately by perturbation from Lemma 4.1 in [6] (which concerns ergodic birth-and-death kernels and is in turn an immediate consequence of [10, Theorem 3.2]. Reversibility of P plays a key role.) Thus the matrix Λ defined at (2.3) is stochastic, as of course is the pure-birth kernel $\widehat{P}$ defined at (1.1). For stochastic P, $\widehat{P}$, and Λ, the identity $\Lambda P^t = \widehat{P}^t \Lambda$ at (2.4) is read as "the semigroups $(P^t)_{t \geq 0}$ and $(\widehat{P}^t)_{t \geq 0}$ are intertwined by the link Λ".

Whenever we have such an intertwining and (to be specific) $X_0 = 0$, Section 2.4 of [2] shows one way to construct explicitly, from X and independent randomness, another Markov chain, say $\widehat{X}$, with kernel $\widehat{P}$ such that

$$\mathcal{L}(X_t \,|\, \widehat{X}_0, \ldots, \widehat{X}_t) = \Lambda(\widehat{X}_t, \cdot) \quad \text{for all } t. \tag{3.1}$$

In particular, since the link Λ is lower triangular [by Lemma 2.1(i)] and $\Lambda(d, d) = 1$ (recall the proof of Lemma 2.3), so that $\Lambda(d, \cdot)$ is unit mass at d, it follows that $\widehat{X}$ and X (and also Y of Theorem 1.2) have the same absorption time T. The discussion in [2] is used in that paper in the case that P is ergodic and $\widehat{P}$ is absorbing, but it applies equally well to any intertwining. Once we have built the pure birth chain $\widehat{X}$, the independent geometric random variables summing to T are simply the waiting times between successive births in $\widehat{X}$.

Here is our construction of $\widehat{X}$; it has the same form (but with $\widehat{\Lambda}$ changed to Λ) as in Section 4.1 of [6], which applied to a different setting, and we repeat it here for the reader's convenience. The chain X starts with $X_0 = 0$ and we set $\widehat{X}_0 = 0$. Inductively, we will have $\Lambda(\widehat{X}_t, X_t) > 0$ (and so $X_t \leq \widehat{X}_t$) at all times t. The value we construct for $\widehat{X}_t$ depends only on the values of $\widehat{X}_{t-1}$ and X_t and independent randomness. Indeed, given $\widehat{X}_{t-1} = \hat{x}$ and $X_t = y$, if $y \leq \hat{x}$ then our construction sets $\widehat{X}_t = \hat{x} + 1$ with probability

$$\frac{\widehat{P}(\hat{x}, \hat{x}+1)\Lambda(\hat{x}+1, y)}{(\widehat{P}\Lambda)(\hat{x}, y)} = \frac{(1 - \theta_{\hat{x}})\Lambda(\hat{x}+1, y)}{\theta_{\hat{x}}\Lambda(\hat{x}, y) + (1 - \theta_{\hat{x}})\Lambda(\hat{x}+1, y)} = \frac{(1 - \theta_{\hat{x}})Q_{\hat{x}+1}(0, y)}{(Q_{\hat{x}}P)(0, y)} \tag{3.2}$$

and $\widehat{X}_t = \hat{x}$ with the complementary probability; if $y = \hat{x} + 1$ (which is the only other possibility, since $y = X_t \leq X_{t-1} + 1 \leq \hat{x} + 1$ by induction), then we set $\widehat{X}_t = \hat{x} + 1$ with certainty.

Remark 3.1. (a) By (3.1) and the lower-triangularity of Λ, our construction satisfies $X_t \leq \widehat{X}_t$ for all t. Thus, among all discrete-time birth-and-death chains on $\{0, \ldots, d\}$ started at 0 and with absorbing state d and given nonnegative eigenvalues $0 \leq \theta_0 \leq \theta_1 \leq \cdots \leq \theta_{d-1} < \theta_d = 1$, the pure-birth "spectral" kernel $\widehat{P}$ defined at (1.1) is stochastically maximal at every epoch t.

(b) In the general setting of Theorem 1.2, we do not know any broad class of examples other than the birth-and-death chains we have just treated for which the eigenvalues are nonnegative real numbers and the spectral polynomials are nonnegative matrices. Nevertheless, the stochastic construction of the preceding paragraph applies verbatim to all such chains.

For continuous-time birth-and-death chains, and more generally for continuous-time skip-free chains with real eigenvalues and generator G having nonnegative spectral poly-

nomials

$$(G + \nu_0 I) \cdots (G + \nu_{k-1} I), \quad k = 0, \ldots, d,$$

there is an analogous construction. It takes a bit more effort to describe than for discrete-time chains, so we refer the reader to Section 5.1 of [6], where again $\widehat{\Lambda}$ there is changed to Λ here, for details. Briefly, if the bivariate chain $(\widehat{X}, X)$ is in state $(\hat{x}, x)$ at a given jump time, then we construct an exponential random variable with rate

$$r = \nu_{\hat{x}} \Lambda(\hat{x} + 1, x) / \Lambda(\hat{x}, x).$$

If X jumps before the lifetime of this exponential expires, then $\widehat{X}$ holds unless X jumps to $\hat{x} + 1$, in which case $\widehat{X}$ also jumps to $\hat{x} + 1$. But if the exponential expires first, then at that expiration time X holds but $\widehat{X}$ jumps to $\hat{x} + 1$.

Remark 3.2. Among all continuous-time birth-and-death chains on $\{0, \ldots, d\}$ started at 0 and with absorbing state d and given eigenvalues $\nu_0 \geq \nu_1 \geq \cdots \geq \nu_{d-1} > \nu_d = 0$ for $-G$, the pure-birth "spectral" chain with birth rate ν_i at state i for $i = 0, \ldots, d$ is stochastically maximal at every epoch t.

4   Proof of Theorem 1.4(b)

In this section we prove Theorem 1.4(b). The proof is quite similar to that of Theorem 1.2, so we shall be brief. Denote the eigenvalues of P by $\theta_0, \ldots, \theta_d$. Since we assume that P is ergodic, precisely one of these, say θ_d, equals 1 and the rest have modulus smaller than 1.

Define the matrices Q_k as at (2.1) [so that (2.2) holds] and define Λ as at (2.3). Then Lemma 2.1 holds for P of Theorem 1.4(b) by the same proof, as do properties (iv)–(v) of Section 1 and the relation (2.4). As we will show next, Lemma 2.3 also holds in the present setting, and then the same argument as in the final paragraph of Section 2 completes the proof of Theorem 1.4(b).

Lemma 4.1. *Let P and X, with fastest strong stationary time T, be as in Theorem 1.4(b) and let $\widehat{P}$ be the bidiagonal matrix defined at (1.1), where $\theta_0, \ldots, \theta_{d-1}$ are the non-unit eigenvalues of P. Then*

$$\mathbf{P}(T \leq t) = \widehat{P}^t(0, d), \quad t = 0, 1, \ldots .$$

Proof. As in the proof of Lemma 2.3, the key is the intertwining relation

$$\Lambda P^t = \widehat{P}^t \Lambda, \quad t = 0, 1, \ldots, \tag{4.1}$$

together with properties (iv)–(v) of Section 1. We consider the $(0, d)$ entries in (4.1).

On the left we have

$$(\Lambda P^t)(0, d) = \sum_k \Lambda(0, k) P^t(k, d) = P^t(0, d) = \mathbf{P}(T \leq t) \, \pi(d), \tag{4.2}$$

where π is the stationary distribution for P. We explain how to obtain the final equality. By the assumption that the time-reversal (call it $\tilde{P}$) of P is stochastically monotone,

the ratio $P^t(0,x)/\pi(x) = \tilde{P}^t(x,0)/\pi(0)$ is minimized by $x = d$, with minimum value (by definition) $1 - s(t)$, where $s(t)$ is the so-called separation from stationarity for X. But, as is well known, a fastest strong stationary time T for X satisfies

$$\mathbf{P}(T \leq t) = 1 - s(t) \quad \text{for all } t.$$

On the right at (4.1) we have

$$(\widehat{P}^t\Lambda)(0,d) = \sum_k \widehat{P}^t(0,k)\Lambda(k,d) = \widehat{P}^t(0,d)\Lambda(d,d). \tag{4.3}$$

Now equate (4.2) and (4.3) and pass to the limit as $t \to \infty$. On the left, $\mathbf{P}(T \leq t) \to 1$ (by ergodicity of X); on the right, $\widehat{P}^t(0,d) \to 1$ by the argument presented in the final paragraph of the proof of Lemma 2.3. So $\Lambda(d,d) = \pi(d)$, and now once again equating (4.2) and (4.3) and canceling the factor $\pi(d) > 0$ gives the desired result. $\square$

Stochastic constructions for Theorem 1.4 have already been given for birth-and-death chains, in discrete time with nonnegative eigenvalues and in continuous time, in Section 4.1 (respectively, Section 5.1) in [6]. The constructions extend verbatim to all skip-free chains with real (and, in discrete time, nonnegative) eigenvalues and nonnegative spectral polynomials.

5 General chains

Our proof in Section 2 of the central Theorem 1.2 for the absorption time of a discrete-time skip-free chain rested on the construction of a matrix Λ having the properties (i)–(iii) listed in Section 1. The question we wish to address in this section are:

(a) Can Theorem 1.2 be extended to general absorbing chains?

(b) Is the spectral-polynomials construction of Λ inevitable? That is, is the matrix Λ uniquely determined by the properties (i)–(iii) listed in Section 1?

(c) If the eigenvalues and spectral polynomials of a general absorbing chain are all nonnegative, can the stochastic construction for Theorem 1.2 be extended?

As we shall see, the answer to each of these questions is **yes**. Question (b) arises naturally because (switching to continuous time and limiting attention to birth-and-death chains for the moment, since that is the setting of [3]) the proofs of Theorem 1.1 both of Diaconis and Miclo [3] and of the present paper rely on construction of a link Λ such that $\Lambda G = \widehat{G}\Lambda$, where $\widehat{G}$ (the analogue of $\widehat{P}$ in continuous time) is the pure-birth "spectral" generator described in our Remark 3.2. The two methods of construction are strikingly different, so it is interesting that the end-product Λ is the same.

All three questions can be addressed in the following setting generalizing that of Theorem 1.2. (We shall subsequently call this "the general setting".) Consider a discrete-time Markov chain X with state space $\{0, \ldots, d\}$ and arbitrary initial distribution m_0 (regarded as a row vector in later calculations) and transition matrix P. Assume that

state d is absorbing and accessible from each other state. (There is no loss of generality in restricting the absorbing set to be a singleton.) Let $\theta_0, \ldots, \theta_{d-1}$ be the d non-unit eigenvalues of P (in fixed but arbitrary order—except where we require the nonnegativity of spectral polynomials, in which case we use nondecreasing order), let $\theta_d = 1$, and let $\widehat{P}$ be defined by (1.1). All square matrices we consider have rows and columns indexed by the state space $\{0, \ldots, d\}$. Let $Q_0, \ldots, Q_d$ be the normalized spectral polynomials defined at (2.1).

Questions (a) and (b) are answered in Section 5.1 (see Lemma 5.1 and Theorem 5.2 respectively); question (c), in Section 5.2.

5.1 Absorption times for general chains and the inevitability of spectral polynomials

Our first result of this subsection demonstrates the inevitable use of spectral polynomials in the construction of Λ, even in the present general setting, provided $\lambda_0 = m_0$.

Lemma 5.1. *The unique matrix Λ (with rows denoted by $\lambda_0, \ldots, \lambda_d$) satisfying the two conditions*

$$m_0 = \lambda_0 \qquad \text{and} \qquad \Lambda P = \widehat{P}\Lambda \tag{5.1}$$

is given by

$$\lambda_i = m_0 Q_i, \quad i = 0, \ldots, d. \tag{5.2}$$

Proof. It is easy to check, as in the proof of Lemma 2.1, that the choice (5.2) satisfies (5.1). Conversely, the ith row of $\Lambda P = \widehat{P}\Lambda$ ($i = 0, \ldots, d-1$) requires

$$\lambda_i P = \theta_i \lambda_i + (1 - \theta_i)\lambda_{i+1}, \quad \text{i.e.,} \quad \lambda_{i+1} = \lambda_i \left[(1 - \theta_i)^{-1}(P - \theta_i I)\right],$$

and so (by induction) (5.1) implies (5.2). $\qquad\square$

Our next result (Theorem 5.2), the main result of this section, greatly generalizes Theorem 1.2 by providing a tidy representation of the absorption-time distribution in the general setting. For the purposes of this result, we use the conventions that an empty sum vanishes and that $\Lambda(-1, d) := 0$ and $\Lambda(d + 1, d) := 1$. We also use the notation

$$a_k := \Lambda(k, d) - \Lambda(k - 1, d), \quad k = 0, \ldots, d + 1. \tag{5.3}$$

Observe that these a_k's sum to unity, and in the proof of Theorem 5.2 we show that a_{d+1} always vanishes. Moreover, if the spectral polynomials happen to be nonnegative, it is easily verified that $\mathbf{a} := (a_0, \ldots, a_{d+1})$ has real and nonnegative entries and so is a probability mass function. Finally, if the eigenvalues θ_i are all real and nonnegative, then for $k = 0, \ldots, d$ we let $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$ denote the convolution of geometric distributions with respective success probabilities $1 - \theta_0, \ldots, 1 - \theta_{k-1}$; we utilize the natural conventions here that this distribution is concentrated at 0 when $k = 0$.

Theorem 5.2. *In the above general setting and notation, in particular with $\widehat{P}$ defined at (1.1), Λ defined as in Lemma 5.1 at (5.2), and a_k defined at (5.3), the absorption*

time T satisfies

$$\mathbf{P}(T \leq t) = \sum_{k=0}^{d} a_k \sum_{j=k}^{d} \widehat{P}^t(0,j), \quad t = 0,1,2,\ldots, \tag{5.4}$$

with probability generating function

$$\mathbf{E}\, u^T = \sum_{k=0}^{d} a_k \prod_{j=0}^{k-1} \left[\frac{(1-\theta_j)u}{1-\theta_j u} \right]. \tag{5.5}$$

In particular, if the eigenvalues θ_i are all nonnegative real numbers and the spectral polynomials in P are all nonnegative matrices, then T is distributed as the $\mathbf{a}$-mixture $\sum_{k=0}^{d} a_k\, \mathcal{G}(\theta_0,\ldots,\theta_{k-1})$ of the convolution distributions $\mathcal{G}(\theta_0,\ldots,\theta_{k-1})$, $k = 0,\ldots,d$.

Proof. Lemma 5.1 is the key. Readily generalizing the proof of Lemma 2.3 and then using summation by parts, one finds

$$\mathbf{P}(T \leq t) = \sum_{j=0}^{d} \widehat{P}^t(0,j)\Lambda(j,d) = \sum_{k=0}^{d} a_k \sum_{j=k}^{d} \widehat{P}^t(0,j).$$

As shown in the proof of Lemma 2.3, $\widehat{P}^t(0,j) \to \delta_{d,j}$ as $t \to \infty$; thus $1 = \sum_{k=0}^{d} a_k$ and so $a_{d+1} = 0$. Equation (5.4) is all that is needed to establish (5.5) when the eigenvalues of P are nonnegative real numbers; in general one can use (2.6) (we omit the routine details). $\quad\square$

Remark 5.3. (a) If the chain is upward skip-free and $m_0 = \delta_0$, then $a_k \equiv \delta_{d,k}$ and Theorem 1.2 is recovered.

(b) The following special case is treated in detail by Miclo [11], albeit in somewhat different fashion. If there exists π satisfying the detailed balance condition $\pi_i p_{ij} = \pi_j p_{ji}$ for all $i,j \in \{0,\ldots,d-1\}$, then by an argument like the one at the beginning of our Section 3 the eigenvalues of P are nonnegative reals and the spectral polynomials are nonnegative. Thus, by Theorem 5.2, the absorption time is distributed as a mixture of convolutions of geometric distributions. Miclo also shows that $a_d > 0$ when the states in $\{0,\ldots,d-1\}$ all communicate with one another.

(c) At least one case other than that of part (b) of this remark is known for which the eigenvalues and spectral polynomials are all nonnegative. If P is upper triangular, then of course the eigenvalues are nonnegative reals, and He and Zhang [7, Appendix A] prove that the spectral polynomials are nonnegative.

(d) A result analogous to Theorem 5.2 holds for continuous-time chains; we omit the details. See Miclo [11, Section 1] for a discussion of connections with the extensive literature on so-called "phase-type" distributions.

(e) A result similar to Theorem 5.2 (with a similar proof) holds for the distribution of the fastest strong stationary time of a general ergodic chain X with general initial distribution m_0 and stationary distribution π, provided $\mathbf{P}(X_t = x)/\pi(x)$ is minimized for every t by the choice $x = d$. A sufficient condition for this is that the time-reversal

of P is stochastically monotone (with respect to the natural linear order on $\{0, \ldots, d\}$) and $m_0(x)/\pi(x)$ is decreasing in x (for example, $m_0 = \delta_0$). The theorem then has the same form as Theorem 5.2, except that now a_k needs to be defined as

$$[\Lambda(k, d) - \lambda(k - 1, d)]/\pi(d).$$

5.2   Stochastic construction via a modified link

We continue to study the general setting. Whenever the eigenvalues and spectral polynomials are all nonnegative, the link Λ of (5.2) provides an intertwining of the semigroups $(P^t)_{t \geq 0}$ and $(\widehat{P}^t)_{t \geq 0}$ [recall (5.1)], and again (as in Section 3) a chain $\widehat{X}$ with kernel $\widehat{P}$ can be constructed such that $\widehat{X}_0 = 0$ and the "sample-path intertwining" relation (3.1) holds. However, unlike for skip-free chains, in the general case although we do have $\Lambda(d, d) = 1$ (this simply restates our earlier observation that $a_{d+1} = 0$) there is no guarantee that the link Λ is lower triangular and thus all we can say with certainty is that the absorption times T for X and $\widehat{T}$ (say) for $\widehat{X}$ satisfy $T \leq \widehat{T}$. We will rectify this situation by modifying the link Λ; this will not contradict the uniqueness of Λ proven in Lemma 5.1, because we will also substitute a different "dual" kernel $\overline{P}$ for the spectral kernel $\widehat{P}$.

To set the stage in the general setting without yet imposing any assumptions about nonnegativity of eigenvalues or spectral polynomials, define the matrix $\overline{\Lambda}$ with rows $\bar{\lambda}_0, \ldots, \bar{\lambda}_d$ by setting

$$\bar{\lambda}_i := \begin{cases} [1 - \lambda_i(d)]^{-1}[\lambda_i - \lambda_i(d)\delta_d] & \text{if } \lambda_i(d) \neq 1 \\ \delta_d & \text{if } \lambda_i(d) = 1, \end{cases} \qquad (i = 0, \ldots, d)$$

where δ_d is the coordinate row vector $(0, 0, \ldots, 0, 1)$, and note that the rows of $\overline{\Lambda}$, like those of Λ, sum to unity. Further, define $\overline{P} := B + R$ to be the sum of the bidiagonal upper triangular matrix B and the matrix R with rank at most 1 vanishing in all columns except for the last defined by

$$b_{ij} := \begin{cases} \theta_i & \text{if } j = i \\ \frac{1 - \lambda_{i+1}(d)}{1 - \lambda_i(d)}(1 - \theta_i) & \text{if } j = i + 1 \\ 0 & \text{otherwise,} \end{cases}$$

and [recalling the notation (5.3)]

$$r_{id} := \left[1 - \frac{1 - \lambda_{i+1}(d)}{1 - \lambda_i(d)}\right](1 - \theta_i) = \frac{a_{i+1}}{1 - \lambda_i(d)}(1 - \theta_i)$$

if $\lambda_i(d) \neq 1$, and by

$$b_{ij} := \delta_{ij}, \qquad r_{id} := 0$$

if $\lambda_i(d) = 1$. Observe that the rows of $\overline{P}$ sum to unity. Finally, let $\overline{m}_0$ be the probability row vector

$$\overline{m}_0 := (1 - m_0(d), 0, \ldots, 0, m_0(d)) = (1 - a_0, 0, \ldots, 0, a_0).$$

The following key fact follows by straightforward calculations from Lemma 5.1.

Lemma 5.4. *In the general setting and the above notation,*

$$m_0 = \overline{m}_0 \overline{\Lambda} \qquad \text{and} \qquad \overline{\Lambda} P = \overline{P}\,\overline{\Lambda}.$$

One can use Lemma 5.4 in place of Lemma 5.1 to give another proof of Theorem 5.2. But much more is possible when the eigenvalues and spectral polynomials of P are all nonnegative, as we shall henceforth assume. In that case we have the following conclusions (with all proofs entirely routine):

- $\overline{\Lambda}$ and $\overline{P}$ are both stochastic, and so the semigroups $(P^t)_{t \geq 0}$ and $(\overline{P}^{\,t})_{t \geq 0}$ are intertwined by the link $\overline{\Lambda}$.

- The set $\overline{A}$ of absorbing states for a chain $\overline{X}$ with kernel $\overline{P}$ satisfies $\overline{A} = \{\overline{d}, \ldots, d\}$, where
$$\overline{d} := \min\{i : \lambda_i(d) = 1\} = \min\{i : a_i = 0\} - 1 \in \{0, \ldots, d\},$$
and $a_i = 0$ if and only if $i \geq \overline{d} + 1$.

- For a $\overline{P}$-chain, from each state in $\{0, \ldots, \overline{d} - 1\}$ the states $\overline{d}$ and d are each accessible but none of the other states in $\overline{A}$ is.

- The construction of Section 2.4 of [2] allows us to build, from X and independent randomness, a chain $\overline{X}$ with initial distribution $\overline{m}$ and kernel $\overline{P}$ such that
$$\mathcal{L}(X_t \mid \overline{X}_0, \ldots, \overline{X}_t) = \overline{\Lambda}(\overline{X}_t, \cdot) \quad \text{for all } t.$$

- The time T to absorption in state d for X is the same (sample-pathwise) as the time to absorption (call it $\overline{T}$) in $\overline{A}$ (i.e., in $\{\overline{d}, d\}$).

- Let L be the largest value reached by $\overline{X}$ prior to absorption, with the convention $L := -1$ if the initial state of $\overline{X}$ is d. Then $\mathbf{P}(L = k - 1) = a_k$ for all $k = 0, \ldots, d$. Further, conditionally given $L = k - 1$, the amounts of time it takes for $\overline{X}$ to move up from 0 to 1, from 1 to 2, $\ldots$, from $k - 2$ to $k - 1$, and from $k - 1$ to $\overline{A}$ are independent geometric random variables with respective success probabilities $1 - \theta_0, 1 - \theta_1, \ldots, 1 - \theta_{k-2}, 1 - \theta_{k-1}$.

Thus in the case of nonnegative eigenvalues and spectral polynomials we have enriched the conclusion of Theorem 5.2 by means of a stochastic construction identifying (i) a random variable (namely, $L + 1$) having probability mass function (a_k); and, conditionally given $L + 1 = k$, (ii) individual geometric random variables whose distributions appear in the convolution $\mathcal{G}(\theta_0, \ldots, \theta_{k-1})$.

Remark 5.5. Recall Remark 5.3. If the conditions described there of detailed balance and complete communication within $\{0, \ldots, d - 1\}$ both hold, then $\overline{d} = d$ and thus $\overline{A} = \{d\}$ is a singleton.

References

[1] Brown, M. and Shao, Y. S. Identifying coefficients in the spectral representation for first passage time distributions. *Probab. Eng. Inform. Sci.* **1** (1987), 69–74.

[2] Diaconis, P. and Fill, J. A. Strong stationary times via a new form of duality. *Ann. Probab.* **18** (1990), 1483–1522.

[3] Diaconis, P. and Miclo, L. On times to quasi-stationarity for birth and death processes. *J. Theoret. Probab.* (2009), to appear.

[4] Fill, J. A. Time to stationarity for a continuous-time Markov chain. *Probab. Eng. Info. Scis.* **5** (1991), 61–76.

[5] Fill, J. A. Strong stationary duality for continuous-time Markov chains. Part I: Theory. *J. Theoret. Probab.* **5** (1992), 45–70.

[6] Fill, J. A. The passage time distribution for a birth-and-death chain: Strong stationary duality gives a first stochastic proof. *J. Theoret. Probab.* (2009), to appear.

[7] He, Q-M. and Zhang, H. Spectral polynomial algorithms for computing bi-diagonal representations for phase type distributions and matrix-exponential distributions. *Stoch. Models* **22** (2006), 289–317.

[8] Horn, R. A. and Johnson, C. R. *Matrix Analysis.* Cambridge University Press, 1985.

[9] Karlin, S. and McGregor, J. Coincidence properties of birth and death processes. *Pacific J. Math.* **9** (1959), 1109–1140.

[10] Micchelli, C. A. and Willoughby, R. A. On functions which preserve the class of Stieltjes matrices. *Lin. Alg. Appl.* **23** (1979), 141–156.

[11] Miclo, L. On absorbtion [sic] times and Dirichlet eigenvalues. Preprint, 2008.